

# Case-Control Studies of Common Alleles and Environmental Factors

Neil Caporaso, Nathaniel Rothman, Sholom Wacholder

**It is clear from descriptive and migration studies that most cancer is environmental in origin. Descriptive, case-control and cohort studies have provided the foundation for our understanding of the environmental component of cancer etiology as well as most major causes of morbidity and mortality. We propose that the same epidemiologic methods that have provided fundamental insight into the etiology of cancer in the general population are optimally suited to study the impact of relatively common polymorphisms on chronic disease incidence. In this article, we describe the role of case-control studies in assessing the effects of genes in disease. Some of the advantages and disadvantages of the case-control design, particularly as an alternative to case-control studies nested in a cohort in the context of the study of complex disease, are described. [Monogr Natl Cancer Inst 1999;26:25-30]**

Population-based epidemiologic studies including descriptive, case-control, and cohort studies have provided the foundation for our understanding of cancer etiology. The accumulated knowledge provided by epidemiologic research carried out during the last 50 years has unequivocally shown that the vast majority of cancer derives from environmental exposures, broadly defined. Descriptive epidemiology reveals striking differences in international cancer incidence rates. Migration studies demonstrate that rates can shift within one generation to the rates of the new country (1). Case-control and cohort epidemiology studies have provided convincing evidence for causal relationships between cancer and smoking, alcohol, viral, drug, radiation, occupational, and lifestyle exposures (2).

There is broad agreement that the relationship between environmental exposures and cancer risk is likely to be modified by genetic factors, since (a) most xenobiotics that enter the body and particularly human carcinogens undergo metabolic processing and (b) wide differences in these capacities between people are often due to heredity (3,4). Many of the specific genes and respective substrates involved are known (Table 1), but the magnitude of the risks and understanding whether gene-environment or gene-gene effects are involved remain controversial. We propose that the same epidemiologic methods that have provided fundamental insight into the etiology of cancer in the general population are optimally suited to evaluate the impact of relatively common polymorphisms on chronic disease incidence as well as the potential interaction between these alleles and the environmental exposures that drive disease risk.

This article highlights the advantages of population-based studies. We do not question the central importance that linkage studies have played in elucidating the genetic origins of family cancer syndromes (5). We note, however, that, for the vast majority of human cancer that exhibits familial tendency but not a simple mendelian inheritance pattern, the tools of classic cancer epidemiology may be better adapted to unraveling the joint ef-

fects of environment and genes. This article highlights the advantages of population-based studies in the investigation of common genetic variation in cancer and, in particular, focuses on case-control studies.

Cohort studies, considered in a companion article (6), and case-control studies are sometimes referred to as "association" studies, implicitly reminding others of the truism that "association does not imply causation." Linkage studies, of course, also use statistical evidence to infer causation. Results from any study design need to be evaluated within the context of the historical (7) and ongoing (8) debate regarding the types of evidence that can advance an association to the level of causation.

Studies of genetic markers in population-based studies were responsible for identifying the relation of HLA-B27 and autoimmune disease, apolipoprotein E in late-onset Alzheimer's disease (9,10), and ABO blood group and gastric cancer (11). In the 1970s and 1980s, this approach was used to study putative associations of cancer with metabolic phenotypes (12-15). By the early 1990s, technology allowed germline DNA to be used in study-specific genetic polymorphisms in relation to particular cancers (16,17) as indicated in Table 1. While the evidence for most of the associations is mixed or sparsely studied, the biologic and epidemiologic data generally support associations between NAT2 (the "slow" acetylation phenotype) (18) and bladder cancer and are at least suggestive for GSTM1 and some smoking related-cancers (19-21). Plausible hypotheses for gene-human cancer/disease associations are the focus of active investigation based on the appreciation that genetic traits can influence the disposition of relevant exposures or mechanistically relevant features on the disease pathway (Table 1).

## CASE-CONTROL STUDIES

The case-control design became widely recognized as an efficient and scientifically sound approach with the studies in the early 1950s linking tobacco smoking and lung cancer (37). Since then, the cancer epidemiology community has accumulated extensive experience with population-based case-control studies. Because of the efficiency and the telescoped time required to complete a case-control study relative to a full cohort study, case-control studies have provided initial clues and been a major, if not the major, source of evidence for many established exposure-disease relationships. They can provide evidence for estimates of the relative and attributable risk and sometimes the absolute risk. In fact, the key weakness of case-control studies of environmental factors, reliance on self-report, does not arise

*Affiliation of authors:* Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, MD.

*Correspondence to:* Neil Caporaso, M.D., National Institutes of Health, Executive Plaza South, Rm. 7116, Bethesda, MD 20892 (e-mail: caporaso@nih.gov).

*See "Note" following "References."*

**Table 1.** Cancer, metabolic polymorphisms, and proposed mechanisms

Cancer/disease	Gene (reference No.)	Proposed basis
Lung	CYP1A1 (22) CYP2E1 (23) CYP2D6 (24) GSTM1 NAT2 EPHX	Phase 1 genes that activate known tobacco carcinogens, including polycyclic aromatic hydrocarbons, <i>N</i> -nitrosamines, and aromatic amines  Phase 2 gene involved in detoxification and elimination of carcinogenic epoxides or aromatic amines derived from tobacco
Nasopharyngeal	GSTM1 (25) CYP2E1 (26)	Detoxification of benzo[a]pyrene or other carcinogens in tobacco smoke Activation of nitrosamines may influence risk
Oral	ADH (27)	Enhanced production of carcinogenic byproducts of alcohol metabolism
Gastric	CYP2E1	Activation of nitrosamine carcinogens
Hepatocellular	EPHX (28) GSTM1	Activation and elimination of aflatoxin B1
Esophageal	CYP2E1 (29)	Activation of nitrosamine carcinogens
Bladder	NAT2 (30)	Decreased elimination of aromatic amines in "slow acetylators"
Breast	CYP1B1 CYP17 NAT2 (31)	Metabolism of estradiol to catechol estrogens Metabolism of steroids Decreased detoxification of aromatic amines in smokers
Colorectal (32)	NAT2	Activation of food-borne carcinogens
Prostate	Androgen receptor (33) SRD5A2 (34)	Alter receptor transactivation with impact on androgen effect Metabolic activation of testosterone to dihydrotestosterone
Non-Hodgkin's lymphoma	CCR5 (35)	In patients with human immunodeficiency virus, chemokine receptor defect alters risk of infection and development of acquired immunodeficiency disease syndrome-related cancer
Renal	CYP1A1 (36)	Phase 1 gene that activates tobacco carcinogens

in studies of constitutional risk factors. Case-control studies conducted in the setting of an established cohort study are described in an accompanying article (6).

## BIOMARKERS OF GENETIC SUSCEPTIBILITY

Modern epidemiologic studies that evaluate genetic factors include DNA-based genotype determinations. When genotype could be determined only by laborious metabolic phenotype determinations involving probe drug administration or complex *in vitro* laboratory studies, investigators encountered methodologic difficulties, including selection bias (i.e., only relatively healthy subjects could undergo phenotyping) and limited power (i.e., difficult laboratory assays or clinical requirements imposed time, cost, and labor constraints on study size). The difficulties of accommodating phenotype assays in field studies generally limited the scope to the characterization of single genes. Today, the genotype information is directly assessed by analysis of germline DNA, obtained from blood or increasingly from other sources (mouthwash rinse, buccal swabs, or stored tissue). Still, the requirement for biospecimens, though less arduous than that imposed during the phenotyping era, continues to have important implications for study design (38,39), measurement error, and sample collection and processing (40).

Beyond issues that derive from the need to obtain biomarkers on study subjects (41), the study of genetic factors imposes further considerations. One issue is that the exploding genetic information and technology will result in multiple and posteriori comparisons. A particular issue that is seemingly unique to associations with genetic markers is the problem of population stratification, considered briefly in the next section.

## POPULATION STRATIFICATION

Population stratification is a consequence of different rates of disease in different ethnic groups; any genetic or environmental

factor whose distribution differs between ethnic groups may appear to be related to disease even though there is no causal relationship. In the simplest situation, if there is a difference in risk between two ethnic groups in the study base, the ethnic groups will be distributed differently between the case and control groups so that any genetic (or other environmental) factor that differs between ethnic groups will tend to be associated with the risk of disease. The worst case for population stratification occurs when two separate populations with widely different allele frequencies and disease rates are admixed as described by Knowler et al. (42) in the study of diabetes and the Gm haplotype in the Pima Indians. An empirical investigation of the extent of bias from the slow acetylation genotype or phenotype and male bladder or female breast cancer found very little bias in a study base of non-Hispanic European-Americans in a multiethnic setting in the United States (Wacholder S et al.: submitted for publication, 1999). We further showed that, when multiple populations are incorporated, as is likely in most U.S. case-control studies, confounding will be minimal. Controlling crudely for ethnicity will often (though not always) control some of the remaining bias. We, therefore, disagree with those who have stated that population stratification undermines the conclusions from population-based studies involving candidate genes so pervasively that familial controls are required to rule out this problem (43,44). The key question is whether a population-based case-control study is so severely affected by this bias that the results are not credible or whether, as we believe, the bias is small and tolerable.

Even if population control subjects have general utility, eliminating residual concerns about population stratification, identifying gene-environment interactions (45), and efficiency considerations may support the use of relative control subjects for a particular study. The relative efficiency of using relatives as control subjects against unmatched population control sub-

jects is considered in another article (46) in this volume. The cost efficiency of selecting an appropriate population control subject and a parent or relative will affect the statistical efficiency as a function of factors such as age, residence of relatives, and the availability of potential rosters. For older control subjects, parents and siblings may be deceased; already identified case subjects will need to be excluded because no eligible control subject is available. Relatives may live far away; extra cost to locate and visit them will be incurred, unless self-administered DNA collection, such as a buccal cell swab, and questionnaire are used. We suspect that, given the added complexity of choosing relative control subjects, the total cost will be higher than that for population control subjects. Unwarranted concern about population stratification, even when it is likely to be trivial, can render the costs of epidemiologic studies prohibitive.

## THE CASE-CONTROL METHOD IN RELATION TO OTHER EPIDEMIOLOGIC DESIGNS

Population-based case-control studies have had a crucial role in unraveling complex human diseases (47-51). Now, however, existing cohort studies that have or are collecting blood samples or other sources of genomic DNA will accrue large numbers of case subjects with common tumors over the coming years. Appropriately, questions are being raised about the utility of carrying out new case-control studies, either population based or hospital based, to study the main effects of common polymorphisms and their interaction with environmental exposures. Designers of a new case-control study will need to show that it offers benefits that cannot be obtained from existing cohorts already under study. We, therefore, raise some considerations when planning to carry out a new case-control study involving genetic factors in contrast to performing nested studies within existing cohorts.

## CONSIDERATIONS BEFORE LAUNCHING A NEW CASE-CONTROL STUDY

### Tumor Incidence

Perhaps the key advantage of case-control studies is the ability to enroll relatively large numbers of cases of the less common tumors. Given the need for sample sizes up to several thousand cases and controls to study gene-environment interaction (52,53), it is only feasible to collect enough cases of the more common tumors in most cohort studies without pooling across cohorts.

### Inclusion of Diverse Population Groups

Case-control studies can focus on enrolling a narrow range of ethnic and racial features, age, or socioeconomic levels that is particularly interesting or important but not adequately represented in existing cohort studies.

### Exposure

**Depth of exposure data.** Case-control studies can collect more detailed and broader information about exposure from both interviews and records than is feasible in a cohort study. This is particularly important when there is concern about a specific type of exposure that is not generally assessed at all or in adequate detail in the typical cohort questionnaire (which usually

focuses on diet and general lifestyle factors). Examples could include occupational and environmental exposures requiring complete occupational and residential histories, respectively. Cohorts have an inherent limitation, in that, having as their aim the study of multiple end points, they can collect less extensive data on exposures relevant to any one particular disease, although the opportunity to return to participants at later time points may partially ameliorate this point. Case-control studies can rapidly respond to and focus on new exposures that are currently of concern for particular tumors, tailoring methods to optimally capture target data. In contrast, cohort studies will have instruments in place that will inevitably lack precision or entirely miss new exposures.

### Retrospective versus prospective exposure determination.

Studies that rely on retrospective exposure assessment that may be affected by disease or its treatment or on questionnaire responses susceptible to rumination by respondents are liable to bias from differential misclassification. Biomarkers (except germline DNA) and responses to questionnaires may change as a consequence of the early disease process or diagnosis itself. We have previously shown a striking drop-off in power to detect gene-environment interactions occurring under some circumstances with even small to moderate exposure misclassification (38,53). Furthermore, even with very rapid case ascertainment, case mortality is a problem for the study of very aggressive tumors in population-based case-control studies. The reliance on prospective exposure assessment preceding the onset of disease protects against differential misclassification. Biochemical measures of exposure are more suitable in cohort settings if the time of collection (i.e., not influenced by preclinical disease) and biologic characteristics (i.e., not influenced by random within-person variation) are appropriate. In addition, case-control studies are usually limited to one retrospective measurement of exposure, whereas cohort studies may conduct repeated measures over time. Although it is feasible to analyze genetic polymorphism in tumor blocks, we have found that the proportion of samples that can be successfully studied is highly variable across studies. Tissue availability may vary by factors related to genetics; i.e., availability of tissue may be limited to "early" surgical cases.

**Gene-environment interactions.** The strong focus on collection of exposure data is well suited to evaluate evidence for interaction between genes and exposure (54). A number of "variant" case-control designs have been proposed to facilitate the study of interaction, particularly the case-only or case-case design (55-57). Subject to certain assumptions (i.e., exposure and gene are independent) and limitations (i.e., the independent effects of gene and exposure cannot be assessed; generally but not always limited to departures from multiplicativity), this approach may be applicable to both case-control studies and nested studies.

### Case and Control Selection

**Ascertainment.** A major concern of case-control studies is proper case and control selection. Proper control subjects are representative of the study base from which the case subjects arise (58,59). Identifying either a random sample from the general population or the source population for case subjects presenting at a particular hospital(s) may be difficult.

**Participation.** Further potential for selection bias occurs if case or control subjects are less likely to participate because of



problems in the collection of biospecimens. Since the source population for cohorts is explicit, selection bias is less of a problem as long as follow-up rates are high (60). Low participation rates in case-control studies and, particularly, refusals related to providing DNA can bias results, especially when case subjects are less likely to participate than control subjects and selection is related to the gene. Low participation rates also threaten the population-based nature of the study, undermining its use for estimating absolute and attributable risks (60). A promising solution to low participation rates for phlebotomy is the less invasive collection of buccal cells (61).

### Collection of Biospecimens

Case-control studies, particularly when hospital-based, offer distinct advantages for the collection of biospecimens. It is not feasible for a large cohort study to reach newly incident cases at a widely dispersed set of hospitals. A case-control study in a narrow geographic region or with the use of only a few hospitals can establish an efficient collection of blood, urine, cells, surgical tissue, and pathologic material along with supporting documentation (medical records). Although rapid contact is not important for germline-derived genetic markers because they are invariant to the presence of disease in the host, case-control studies also have an advantage if fresh tumor tissue is a study requirement. Resources for collection, processing, and storage are often conveniently available around the hospital setting. Often, pretreatment specimens, critical for evaluation of biologic markers that could be affected by chemotherapy or radiation therapy, can be obtained. Furthermore, case-control studies can generally collect larger quantities and types of biologic samples and process them in ways (i.e., cryopreservation of lymphocytes and Epstein-Barr virus transformation to ensure large quantities of DNA) that are more sophisticated than is feasible in cohort studies. This offers the potential for conducting functional assays such as mutagen sensitivity (62), which in general are not methodologically feasible in cohort settings.

### Single Disease

Case-control studies are generally limited to one disease outcome but are less constrained by the rarity of the disease, whereas cohort studies (including full cohort, nested case-control, or case-cohort studies) may identify multiple disease end points. The focus of a case-control study on one disease entity permits greater time and resources to be devoted to the collection and documentation of disease information (i.e., pathology, staging, and tissue), providing a richer information base for study with the benefits of a potential reduction in misclassification and a deeper analysis. Obtaining disease-related data in cohorts entails mounting an effort that is generally less efficient and more costly. The advantage of cohort studies' ability to examine multiple outcomes may be somewhat limited by resources and logistics, limited exposure information, the diverse approaches to documenting disease incidence or mortality, and the rarity of some outcomes.

Costs for a series of case-control studies of different cancer sites can sometimes be reduced by sharing a single control group. When different diseases require different exposures, the partial questionnaire design may offer a reduction in the burden to respondents, thereby potentially increasing participation (63). Even if these options are not feasible, using the same infrastructure for control selection for repeated studies can reduce costs.

## Resources and Infrastructure

The vastly greater size of cohorts large enough to generate a substantial sample of diseased subjects and the time period required for the cohort to mature mean that a substantially greater initial investment is required to establish the cohort. For cohort studies that incorporate biologic materials, the infrastructure to support biospecimens' databases, freezers, and processing requires a correspondingly greater effort and cost. While all studies with biospecimens must consider the risk of untoward events (i.e., freezer failure), the anticipated long useful life of the samples from cohorts requires special emphasis on quality control and security issues (i.e., backup generators, monitoring, distributing samples among different freezers, etc.). In the next few years, however, the "cost per case" for studies fielded from a cohort will offer economies in comparison to fielding a new case-control study (64).

## FUTURE TRENDS

The primacy of exposure in cancer etiology and the recognition that cancer is manifestly hereditary in certain families, but consistently a genetic disease on the cellular level mandate an interdisciplinary approach to elucidating their joint role in cancer. Family studies have proven critical to identifying high-penetrance genes. Mechanistic work and animal work have provided extraordinary insight into our fundamental understanding of cancer. Unraveling the complex origin of cancer that results in the vast burden of cancer mortality will require population-based studies that can uniquely determine the public health impact of a putative risk factor, genetic or otherwise.

Population studies that include tissue collection and exposure assessment will allow opportunities for diverse hypotheses testing involving, e.g., relations between germline and tumor mutations, exposure relationships in questionnaire-based and molecular approaches, and exploitation of the developing human genome map for both population and family-based studies. There will be opportunities to explore for unique and newly hypothesized etiologic agents in stored material, to correlate gene and exposure markers with tumor mutations, and generally to explore mechanistic hypotheses. The next generation of studies that will strive to understand the interplay between environmental and genetic factors for cancer risk in the general population will be substantially larger than previous biologically based case-control studies in order to have adequate power to detect interactions between environmental and genetic factors (38). Generally, larger study sizes should also permit new approaches requiring pooled DNA samples that exploit the increasingly detailed genetic map to conduct gene searches, i.e., linkage disequilibrium mapping (65) with the use of unrelated control subjects. Less invasive methods of genomic DNA collection, such as buccal cell collections (61), will find increasing application and will benefit all study approaches involving biospecimens.

In the future, both case-control and cohort studies will have crucial and complementary roles in the investigation of genetic factors and their interaction with environmental exposures in the general population. As the number of cases with common tumors accumulate in existing cohorts over the coming years, nested case-control studies will have an increasingly important role, with special advantages deriving from large size, multiple outcomes (both cancer and non-cancer), and prospective expo-

sure assessment. Traditional case-control studies and their variants will be central for focused investigations of single cancers, where detailed exposure and intensive biologic sample collection are deemed necessary to integrated investigations. The argument for the case-control method will be particularly compelling when new or changing exposures must be quickly evaluated, when populations underrepresented in cohort studies must be rapidly investigated, or when less common tumors are the focus of study.

## REFERENCES

- (1) Thomas DB, Karagas MR. Migrant studies. In: Schottenfeld D, Fraumeni JF Jr. *Cancer epidemiology and prevention*. 2<sup>nd</sup> ed. Oxford (U.K.): Oxford University Press, 1996. p. 236-54.
- (2) Schottenfeld D, Fraumeni JF Jr. *Cancer epidemiology and prevention*. 2<sup>nd</sup> ed. Oxford (U.K.): Oxford University Press; 1996.
- (3) Frame LT, Ambrosone CB, Kadlubar FF, Lang NP. Host-environment interactions that affect variability in human cancer susceptibility. In: Neumann DA, Kimmel CA, editors. *Human variability in response to chemical exposures*. Washington (DC): International Life Sciences Institute; 1998. p. 165-204.
- (4) Lang M, Pelkonen O. Metabolism of xenobiotics and chemical carcinogenesis. In: *Metabolic polymorphisms and susceptibility to cancer*. IARC Sci Publ 1999;148:13-22.
- (5) Lindor NM, Greene MH, the Mayo Familial Cancer Program. The concise handbook of family cancer syndromes. *J Natl Cancer Inst* 1998;90:1039-71.
- (6) Langholz B, Rothman N, Wacholder S, Thomas DC. Cohort studies for characterizing measured genes. *Monogr Natl Cancer Inst* 1999;26:39-42.
- (7) Hill AB. The environment and disease: association or causation. *Proc R Soc Med* 1965;58:295-300.
- (8) Rothman KJ. *Modern epidemiology*. 2<sup>nd</sup> ed. Boston (MA): Little Brown & Co.; 1998.
- (9) Kukull WA, Schellenberg GD, Bowen JD, McCormick WC, Yu CE, Teri L, et al. Apolipoprotein E in Alzheimer's disease risk and case detection: a case-control study. *J Clin Epidemiol* 1996;49:1143-8.
- (10) Sanders AM, Strittmatter WJ, Schmechel D, George-Hyslop S, Pericak-Vance MA, Joo SH, et al. Association of apolipoprotein E allele e4 with late onset familial and sporadic Alzheimer's disease. *Neurology* 1993;43:1467-72.
- (11) Aird I, Bentall HH. A relationship between cancer of the stomach and the ABO blood groups. *Br Med J* 1953;1:799-801.
- (12) Kellermann G, Shaw CR, Luyten-Kellermann M. Aryl hydrocarbon hydroxylase inducibility and bronchogenic carcinoma. *N Engl J Med* 1973;289:934-7.
- (13) Ayesh R, Idle JR, Richie JC, Crothers MJ, Hetzel MR. Metabolic oxidation phenotypes as markers for susceptibility to lung cancer. *Nature* 1984;312:169-70.
- (14) Seidegard J, Pero RW, Miller D, Beattie EJ. A glutathione transferase in human leukocytes as a marker for susceptibility to lung cancer. *Carcinogenesis* 1986;7:751-3.
- (15) Lower GM, Nilsson T, Nelson CE, Wolf H, Gamsky TE, Bryan GT. *N*-Acetyltransferase phenotype and risk in urinary bladder cancer: approaches in molecular epidemiology. Preliminary results in Sweden and Denmark. *Environ Health Perspect* 1979;29:71-9.
- (16) Nebert DW, Ingelman-Sundberg M, Daly AK. Genetic epidemiology of environmental toxicity and cancer susceptibility: human allelic polymorphisms in drug metabolizing genes, their functional importance, and nomenclature issues. *Drug Metab Rev* 1999;31:467-82.
- (17) Caporaso N, Goldstein A. Issues involving biomarkers in the study of the genetics of human cancer. *IARC Sci Publ* 1997;142:237-50.
- (18) Marcus PM, Vineis P, Rothman N. NAT2 slow acetylation and bladder cancer risk: a meta-analysis of 22 case-control studies conducted in the general population. *Pharmacogenetics*. In press 2000.
- (19) Rebbeck TR. Molecular epidemiology of the human glutathione *S*-transferase genotypes GSTM1 and GSTT1 in cancer susceptibility. *Cancer Epidemiol Biomarkers Prev* 1997;6:733-43.
- (20) Houlston RS. Glutathione *S*-transferase M1 status and lung cancer risk: a meta-analysis. *Cancer Epidemiol Biomarkers Prev* 1999;8:675-83.
- (21) D'Errico A, Taioli E, Chen X, Vineis P. Genetic metabolic polymorphisms and the risk of cancer: a review of the literature. *Biomarkers* 1996;1:149-73.
- (22) Kawajiri K. CYP1A1. *IARC Sci Publ* 1999;148:159-72.
- (23) Wu XF, Amos CI, Kemp BL, Shi HH, Jiang H, Wan Y, et al. Cytochrome P450 2E1 *Dra*I polymorphisms in lung cancer in minority populations. *Cancer Epidemiol Biomarkers Prev* 1998;7:13-8.
- (24) Shaw GL, Falk JN, Weiffenbach B, Nesbitt JC, Pass HI, Caporaso NE, et al. Genetic polymorphism of CYP2D6 and lung cancer risk. *Cancer Epidemiol Biomarkers Prev* 1998;7:215-9.
- (25) Nazar-Stewart V, Vaughan TL, Burt RD, Chen C, Berwick M, Swanson GM. Glutathione *S*-transferase M1 and susceptibility to nasopharyngeal carcinoma. *Cancer Epidemiol Biomarkers Prev* 1999;8:547-51.
- (26) Hildesheim A, Anderson LM, Chen CJ, Cheng YJ, Brinton LA, Daly AK, et al. CYP2E1 genetic polymorphisms and risk of nasopharyngeal carcinoma in Taiwan. *J Natl Cancer Inst* 1997;89:1207-12.
- (27) Harty LC, Caporaso NE, Hayes RB, Winn DM, Bravo-Otero E, Blot WJ, et al. Alcohol dehydrogenase 3 genotype and risk of oral cavity and pharyngeal cancers. *J Natl Cancer Inst* 1997;89:1698-705.
- (28) McGlynn KA, Rosvold EA, Lustbader ED, Hu Y, Clapper ML, Zhou T, et al. Susceptibility to hepatocellular carcinoma is associated with genetic variation in the enzymatic detoxification of aflatoxin B1. *Proc Natl Acad Sci U S A* 1995;92:2384-7.
- (29) Lin DX, Tang YM, Peng Q, Lu SX, Ambrosone CB, Kadlubar FF. Susceptibility to esophageal cancer and genetic polymorphisms in glutathione *S*-transferases T1, P1, and M1 and cytochrome P450 2E1. *Cancer Biomarkers Epidemiol Prev* 1998;7:1013-8.
- (30) Marcus PM, Hayes RB, Vineis P, Garcia-Closas M, Caporaso NE, Rothman N. NAT2 slow acetylation and cigarette smoking: a meta-analysis of their interactive effect on bladder cancer risk among 1,986 cases [abstract]. *Proc Am Assoc Cancer Res* 1999;40:212.
- (31) Ambrosone CB, Freudenheim JL, Graham S, Marshall JR, Vena JE, Brasure JR, et al. Cigarette smoking, *N*-acetyltransferase 2 genetic polymorphism and breast cancer risk. *JAMA* 1996;276:1494-501.
- (32) Kampman E, Slattery ML, Bigler J, Leppert M, Samowitz W, Caan BJ, et al. Meat consumption, genetic susceptibility, and colon cancer risk: a United States multicenter case-control study. *Cancer Epidemiol Biomarkers Prev* 1999;8:15-24.
- (33) Irvine RA, Yu MC, Ross RK, Coetze GA. The CAG and GGC microsatellites of the androgen receptor gene are in linkage disequilibrium in men with prostate cancer. *Cancer Res* 1995;55:1937-40.
- (34) Makridakis NM, Ross RK, Pike MC, Crocitto LE, Kolonel LN, Pearce CL, et al. Association of mis-sense substitution in SRD5A2 gene with prostate cancer in African-American and Hispanic men in Los Angeles, USA. *Lancet* 1999;354:975-8.
- (35) Dean M, Jacobson LP, McFarlane G, Margolick JD, Jenkins FJ, Howard OM, et al. Reduced risk of AIDS lymphoma in individuals heterozygous for the CCR5-delta 32 mutation. *Cancer Res* 1999;59:3561-4.
- (36) Longueux S, Delomenie C, Gallou C, Mejean A, Vincent-Viry M, Bouvier R, et al. Candidate genetic modifiers of individual susceptibility to renal cell carcinoma: a study of polymorphic human xenobiotic-metabolizing enzymes. *Cancer Res* 1999;59:2903-8.
- (37) Doll R, Hill AB. A study of the aetiology of carcinoma of the lung. *BMJ* 1952;2:1271-86.
- (38) Rothman N, Garcia-Closas M, Stewart WJ, Lubin J. Impact of misclassification in studies of gene-environment interactions. *IARC Sci Publ* 1999;148:89-96.
- (39) Pearce N, Boffetta P. General issues of study design and analysis in the use of biomarkers in cancer epidemiology. *IARC Sci Publ* 1997;142:47-58.
- (40) Landi MT, Caporaso N. Issues in the collection, processing and storage of biospecimens. *IARC Sci Publ* 1997;142:223-6.
- (41) Schulte PA, Perera FP. Transitional studies. *IARC Sci Publ* 1997;142:19-29.
- (42) Knowler WC, Williams RC, Pettitt DJ, Steinberg AG. Gm and type 2 diabetes mellitus: an association in American Indians with genetic admixture. *Am J Hum Genet* 1988;43:5205-26.
- (43) Altshuler D, Kruglyak L, Lander E. Genetic polymorphism and disease. *N Engl J Med* 1998;338:1626.
- (44) Lander ES, Schork NJ. Genetic dissection of complex traits. *Science* 1994;265:2037-48.

- (45) Witte JS, Gauderman WJ, Thomas DC. Asymptotic bias and efficiency in case-control studies of candidate genes and gene-environment interactions: basic family designs. *Am J Epidemiol* 1999;149:693-705.
- (46) Gauderman WJ, Witte JS, Thomas DC. Family-based association studies. *Monogr Natl Cancer Inst* 1999;26:31-7.
- (47) Rothman N, Stewart W, Schulte PA. Incorporating biomarkers into cancer epidemiology: a matrix of biomarker and study design categories. *Cancer Epidemiol Biomarkers Prev* 1995;4:301-11.
- (48) Khoury MJ, Beaty TH. Applications of the case-control method in genetic epidemiology. *Epidemiol Rev* 1994;16:134-50.
- (49) Caporaso N, Goldstein A. Cancer genes: single and susceptibility: exposing the difference. *Pharmacogenetics* 1995;5:59-63.
- (50) Risch N, Merikangas K. The future of genetic studies of complex human diseases. *Science* 1996;273:1516-7.
- (51) Schork NJ, Cardon LR, Xu X. The future of genetic epidemiology. *Trends Genet* 1998;14:266-72.
- (52) Garcia-Closas M, Lubin J. Power and sample size calculations in case-control studies of gene-environment interactions. *Am J Epidemiol* 1999;149:689-92.
- (53) Garcia-Closas M, Rothman N, Lubin J. Misclassification in case-control studies of gene-environment interactions: assessment of bias and sample size. *Cancer Epidemiol Biomarkers Prev*. In press 1999.
- (54) Khoury MJ. Genetic epidemiology and the future of disease prevention and public health. *Epidemiol Rev* 1997;19:175-80.
- (55) Piegorsch WW, Weinberg CR, Taylor JA. Non-hierarchical logistic models and case-only designs for assessing susceptibility in population case-control studies. *Stat Med* 1994;13:153-62.
- (56) Yang Q, Khoury MJ. Evolving methods in genetic epidemiology. Gene-environment interaction in epidemiological research. *Epidemiol Rev* 1997;19:33-43.
- (57) Khoury MJ, Yang Q. The future of genetic studies of complex human diseases: an epidemiologic perspective. *Epidemiology* 1998;9:350-4.
- (58) Wacholder S, McLaughlin JK, Silverman DT, Mandel JS. Selection of controls in case-control studies. *Am J Epidemiol* 1992;135:1019-50.
- (59) Breslow NE, Day NE. Statistical methods in cancer research, vol 1. The analysis of case-control studies. *IARC Sci Publ* 1980;32:84-119.
- (60) Hunter DJ. Methodological issues in the use of biological markers in cancer epidemiology: cohort studies. *IARC Sci Publ* 1997;142:39-46.
- (61) Lum A, Le Marchand L. A simple mouthwash method for obtaining genomic DNA in molecular epidemiologic studies. *Cancer Epidemiol Biomarkers Prev* 1998;7:719-24.
- (62) Wu X, Gu J, Hong WK, Lee JJ, Amos CI, Jiang H, et al. Benzo[a]pyrene diol epoxide and bleomycin sensitivity to cancer of upper aerodigestive tract. *J Natl Cancer Inst* 1998;90:1393-9.
- (63) Wacholder S, Carroll RJ, Pee D, Gail MH. The partial design for case-control studies. *Stat Med* 1994;13:623-34.
- (64) Potter JD. Logistics and design issues in the use of biological samples in observational epidemiology. *IARC Sci Publ* 1997;142:31-7.
- (65) Barcellos LF, Klitz W, Field LL, Tobias R, Bowcock AM, Wilson R, et al. Association mapping of disease loci, by use of a pooled DNA genomic screen. *Am J Hum Genet* 1997;61:734-47.

## NOTE

We thank Alisa Goldstein for thoughtful review of this work.